

Convergence analysis of the Gibbs sampler for Bayesian general linear mixed models with improper priors

Jorge Carlos Román and James P. Hobert

Department of Statistics

University of Florida

November 2011

Abstract

Bayesian analysis of data from the general linear mixed model is challenging because any nontrivial prior leads to an intractable posterior density. However, if a conditionally conjugate prior density is adopted, then there is a simple Gibbs sampler that can be employed to explore the posterior density. A popular default among the conditionally conjugate priors is an improper prior that takes a product form with a flat prior on the regression parameter, and so-called power priors on each of the variance components. In this paper, a convergence rate analysis of the corresponding Gibbs sampler is undertaken. The main result is a simple, easily-checked sufficient condition for geometric ergodicity of the Gibbs Markov chain. This result is close to the best possible result in the sense that the sufficient condition is only slightly stronger than what is required to ensure posterior propriety. The theory developed in this paper is extremely important from a practical standpoint because it guarantees the existence of central limit theorems that allow for the computation of valid asymptotic standard errors for the estimates computed using the Gibbs sampler.

AMS 2000 subject classifications. Primary 60J27; secondary 62F15

Abbreviated title. Gibbs Sampling for Mixed Models

Key words and phrases. Geometric drift condition, Geometric ergodicity, Markov chain, Monte Carlo, Posterior propriety

1 Introduction

The general linear mixed model (GLMM) takes the form

$$Y = X\beta + Zu + e ,$$

where Y is an $N \times 1$ data vector, X and Z are known matrices with dimensions $N \times p$ and $N \times q$, respectively, β is an unknown $p \times 1$ regression coefficient, u is a random vector whose elements represent the various levels of the random factors in the model, and $e \sim N_N(0, \sigma_e^2 I)$. The random vectors e and u are assumed to be independent. Suppose there are r random factors in the model. Then u and Z are partitioned accordingly as $u = (u_1^T u_2^T \cdots u_r^T)^T$ and $Z = (Z_1 Z_2 \cdots Z_r)$, where u_i is $q_i \times 1$, Z_i is $N \times q_i$, and $q_1 + \cdots + q_r = q$. Then

$$Zu = \sum_{i=1}^r Z_i u_i ,$$

and it is assumed that $u \sim N_q(0, D)$, where $D = \oplus_{i=1}^r \sigma_{u_i}^2 I_{q_i}$. For background on this model, which is sometimes called the *variance components model*, see Searle et al. (1992).

A Bayesian version of the GLMM can be assembled by specifying a prior distribution for the unknown parameters β and σ^2 , where $\sigma^2 = (\sigma_e^2 \sigma_{u_1}^2 \cdots \sigma_{u_r}^2)^T$ denotes the vector of variance components. A popular choice is the proper (conditionally) conjugate prior that takes β to be multivariate normal, and takes each of the variance components to be inverted gamma. In situations where there is little prior information, the hyperparameters of this proper prior are often set to extreme values as this is thought to yield a “non-informative” prior. Unfortunately, these extreme proper priors approximate improper priors that correspond to improper posteriors, and this results in various forms of instability. This problem has led several authors, including Daniels (1999) and Gelman (2006), to discourage the use of such extreme proper priors, and to recommend alternative default priors that are improper, but lead to proper posteriors. Consider, for example, the one-way random effects model given by

$$Y_{ij} = \beta_0 + u_i + e_{ij} ,$$

where $i = 1, \dots, c$, $j = 1, \dots, n_i$, the u_i s are iid $N(0, \sigma_1^2)$, and the e_{ij} s, which are independent of the u_i s, are iid $N(0, \sigma_e^2)$. This is the so-called “non-centered” version of the one-way model. It is an important special case of our GLMM. (In the alternative “centered” parameterization, the parameter β_0 does not appear in the model equation, but rather as the mean of the u_i s.) The standard

diffuse prior for this model, which is among those recommended by Gelman (2006), has density $1/(\sigma_e^2 \sqrt{\sigma_1^2})$. In fact, many of the improper priors for the GLMM that have been suggested and studied in the literature take the form of a reciprocal of a product of polynomials in the variance components. One obvious reason for using such a prior is that the resulting posterior has conditional densities with standard forms, and this facilitates the use of the Gibbs sampler.

In this paper, we consider the following parametric family of priors for (β, σ^2) :

$$p(\beta, \sigma^2; a, b) = (\sigma_e^2)^{-(a_e+1)} e^{-b_e/\sigma_e^2} \times \left[\prod_{i=1}^r (\sigma_{u_i}^2)^{-(a_i+1)} e^{-b_i/\sigma_{u_i}^2} \right], \quad (1)$$

where $a = (a_e, a_1, \dots, a_r)$ and $b = (b_e, b_1, \dots, b_r)$ are fixed hyperparameters. By taking b to be 0, we can recover the reciprocal polynomial priors described above. Note that β does not appear on the right-hand side of (1); that is, we are using a so-called flat prior for β . Consequently, even if all the elements of a and b are strictly positive, so that every variance component gets a proper prior, the overall prior remains improper. There have been several studies concerning posterior propriety in this context, and the most general of these was done by Sun et al. (2001). We state their main result here so that it can be used in a comparison later in this section. Define $\theta = (\beta^T \ u^T)^T$ and $W = (X \ Z)$, so that $W\theta = X\beta + Zu$. Let y denote the observed data, and let $\phi_d(x; \mu, \Sigma)$ denote the $N_d(\mu, \Sigma)$ density evaluated at the vector x . By definition, the posterior density is proper if

$$m(y) := \int_{\mathbb{R}_+^{r+1}} \int_{\mathbb{R}^{p+q}} \pi^*(\theta, \sigma^2 | y) d\theta d\sigma^2 < \infty,$$

where

$$\pi^*(\theta, \sigma^2 | y) = \phi_N(y; W\theta, \sigma_e^2 I) \phi_q(u; 0, D) p(\beta, \sigma^2; a, b).$$

The following result provides sufficient (and nearly necessary) conditions for propriety.

Theorem 1. [Sun, Tsutakawa & He (2001)] Assume that $\text{rank}(X) = p$ and let $t = \text{rank}(Z^T(I - X(X^T X)^{-1} X^T)Z)$. If the following four conditions hold, then $m(y) < \infty$.

(A) For each $i \in \{1, 2, \dots, r\}$, one of the following holds:

$$(A1) \ a_i < b_i = 0; \quad (A2) \ b_i > 0$$

(B) For each $i \in \{1, 2, \dots, r\}$, $q_i + 2a_i > q - t$

(C) $N + 2a_e > p - 2 \sum_{i=1}^r a_i I_{(-\infty, 0)}(a_i)$

$$(D) \quad 2b_e + \|(I - W(W^T W)^{-1} W^T) y\|^2 > 0$$

When $m(y) < \infty$, the posterior density is well defined (i.e. proper) and is given by $\pi(\theta, \sigma^2|y) = \pi^*(\theta, \sigma^2|y)/m(y)$. It is well known that $\pi(\theta, \sigma^2|y)$ is intractable in the sense that posterior expectations cannot be computed in closed form, nor even by classical Monte Carlo methods. However, there is a simple two-step Gibbs sampler that can be used to approximate the intractable posterior expectations. This Gibbs sampler simulates a Markov chain, $\{(\theta_n, \sigma_n^2)\}_{n=0}^\infty$, that lives on $X = \mathbb{R}^{p+q} \times \mathbb{R}_+^{r+1}$, and has invariant density $\pi(\theta, \sigma^2|y)$. If the current state of the chain is (θ_n, σ_n^2) , then the next state, $(\theta_{n+1}, \sigma_{n+1}^2)$, is simulated using the usual two steps. Indeed, we draw θ_{n+1} from $\pi(\theta|\sigma_n^2, y)$, which is a $(p+q)$ -dimensional multivariate normal density, and then we draw σ_{n+1}^2 from $\pi(\sigma^2|\theta_{n+1}, y)$, which is a product of $r+1$ univariate inverted gamma densities. The exact forms of these conditional densities are given in Section 2.

Because the Gibbs Markov chain is Harris ergodic (see Section 2), we can use it to construct consistent estimates of intractable posterior expectations. For $k > 0$, let $L_k(\pi)$ denote the set of functions $g : \mathbb{R}^{p+q} \times \mathbb{R}_+^{r+1} \rightarrow \mathbb{R}$ such that

$$E_\pi |g|^k := \int_{\mathbb{R}_+^{r+1}} \int_{\mathbb{R}^{p+q}} |g(\theta, \sigma^2)|^k \pi(\theta, \sigma^2|y) d\theta d\sigma^2 < \infty.$$

If $g \in L_1(\pi)$, then the average $\bar{g}_m := \frac{1}{m} \sum_{i=0}^{m-1} g(\theta_i, \sigma_i^2)$ is a strongly consistent estimator of $E_\pi g$, no matter how the chain is started. Of course, in practice, an estimator is only useful if it is possible to compute an associated standard error. All available methods of computing a valid asymptotic standard error for \bar{g}_m are based on the existence of a central limit theorem (CLT) for \bar{g}_m (Flegal et al., 2008; Jones et al., 2006). Unfortunately, even if $g \in L_k(\pi)$ for all $k > 0$, Harris ergodicity is not enough to guarantee the existence of such a CLT for \bar{g}_m (see, e.g., Roberts and Rosenthal, 1998, 2004). The standard method of establishing the existence of CLTs is to prove that the underlying Markov chain converges at a geometric rate.

Let $\mathcal{B}(X)$ denote the Borel sets in X , and let $P^n : X \times \mathcal{B}(X) \rightarrow [0, 1]$ denote the n -step Markov transition function of the Gibbs Markov chain. That is, $P^n((\theta, \sigma^2), A)$ is the probability that $(\theta_n, \sigma_n^2) \in A$, given that the chain is started at $(\theta_0, \sigma_0^2) = (\theta, \sigma^2)$. Also, let $\Pi(\cdot)$ denote the posterior distribution. The chain is called geometrically ergodic if there exist a function $M : X \rightarrow [0, \infty)$ and a constant $\varrho \in [0, 1)$ such that, for all $(\theta, \sigma^2) \in X$ and all $n = 0, 1, \dots$, we have

$$\|P^n((\theta, \sigma^2), \cdot) - \Pi(\cdot)\|_{\text{TV}} \leq M(\theta, \sigma^2) \varrho^n,$$

where $\|\cdot\|_{\text{TV}}$ denotes the total variation norm. The relationship between geometric convergence and CLTs is simple: If the chain is geometrically ergodic and $E_{\pi}|g|^{2+\delta} < \infty$ for some $\delta > 0$, then \bar{g}_m satisfies a CLT. Our main result (Theorem 2 in Section 3) provides conditions under which the Gibbs Markov chain is geometrically ergodic. Checking these conditions may require some numerical work. The following corollary to Theorem 2 is weaker, but easier to apply.

Corollary 1. *Assume that $\text{rank}(X) = p$. If the following four conditions hold, then the Gibbs Markov chain is geometrically ergodic.*

(A) *For each $i \in \{1, 2, \dots, r\}$, one of the following holds:*

$$(A1) \ a_i < b_i = 0 ; \quad (A2) \ b_i > 0$$

(B') *For each $i \in \{1, 2, \dots, r\}$, $q_i + 2a_i > q - t + 2$*

(C') *$N + 2a_e > p + t + 2$*

(D) *$2b_e + \|(I - W(W^T W)^{-1} W^T) y\|^2 > 0$*

In cases where the posterior density is improper, it is sometimes still possible to run the Gibbs sampler, but the corresponding Markov chains cannot be geometrically ergodic (see Section 2). Therefore, the best we could hope for is that the Gibbs Markov chain is geometrically ergodic whenever the posterior is proper. With this in mind, note that the conditions of Corollary 1 are very close to the conditions for propriety given in Theorem 1. In fact, the former imply the latter. To see this, note that (B') clearly implies (B), which, in turn, implies that

$$\sum_{i=1}^r (q_i + 2a_i) I_{(-\infty, 0)}(a_i) > q - t .$$

Now since $q = q_1 + \dots + q_r$, we have

$$t > q - \sum_{i=1}^r (q_i + 2a_i) I_{(-\infty, 0)}(a_i) \geq -2 \sum_{i=1}^r a_i I_{(-\infty, 0)}(a_i) ,$$

and it follows that

$$p + t + 2 > p + t > p - 2 \sum_{i=1}^r a_i I_{(-\infty, 0)}(a_i) .$$

Hence, (C') implies (C). The strong similarity between Theorem 1 and Corollary 1 might lead the reader to believe that the proofs of our results rely somehow on Theorem 1. This is not the

case. Indeed, the conditions of Corollary 1 arise completely independently. In fact, we do not even assume propriety before embarking on our convergence rate analysis (see Section 3).

Analogues of Corollary 1 for the GLMM with *proper* priors can be found in Johnson and Jones (2010) and Román (2012). The proper and improper cases are similar in the sense that geometric ergodicity is established via geometric drift conditions in both cases. However, the drift conditions are quite disparate, and the analysis required in the improper case is substantially more demanding. In fact, the only other existing results on geometric convergence of Gibbs samplers for linear mixed models with *improper* priors are those of Tan and Hobert (2009), who considered the centered version of the one-way random effects model. Since the centered model is *not* a special case of our GLMM, their results are not special cases of ours. On the surface, the difference between the centered and non-centered models may seem inconsequential, but it is well known that MCMC algorithms corresponding to models that differ only by minor reparameterizations can have radically different convergence rates (see, e.g. Gelfand et al., 1995; Papaspiliopoulos et al., 2007; Yu and Meng, 2011). Finally, we note that the linear models considered by Papaspiliopoulos and Roberts (2008) are substantively different from ours because these authors assume that the variance components are known.

The remainder of this paper is organized as follows. Section 2 contains a formal definition of the Gibbs Markov chain. The main convergence result is stated and proven in Section 3. Finally, Section 4 concerns an interesting technical issue related to the use of improper priors. An oversight by Tan and Hobert (2009) regarding this technicality led to an error in the proof of their main result. However, it is shown in Section 4 that their proof is easily repaired (using the results developed herein), and that their result remains correct as stated in their paper.

2 The Gibbs Sampler

In this section, we formally define the Gibbs sampler, and state some of its properties. Recall that $\theta = (\beta^T \ u^T)^T$ and $\sigma^2 = (\sigma_e^2 \ \sigma_{u_1}^2 \ \cdots \ \sigma_{u_r}^2)^T$. Suppose that

$$\int_{\mathbb{R}^{p+q}} \pi^*(\theta, \sigma^2 | y) d\theta < \infty \quad (2)$$

for all σ^2 outside a set of measure zero in \mathbb{R}_+^{r+1} , and that

$$\int_{\mathbb{R}_+^{r+1}} \pi^*(\theta, \sigma^2 | y) d\sigma^2 < \infty \quad (3)$$

for all θ outside a set of measure zero in \mathbb{R}^{p+q} . These two integrability conditions are necessary, but not sufficient, for posterior propriety. When they hold, we can define conditional densities as follows:

$$\pi(\theta|\sigma^2, y) = \frac{\pi^*(\theta, \sigma^2|y)}{\int_{\mathbb{R}^{p+q}} \pi^*(\theta, \sigma^2|y) d\theta} \quad \text{and} \quad \pi(\sigma^2|\theta, y) = \frac{\pi^*(\theta, \sigma^2|y)}{\int_{\mathbb{R}_+^{r+1}} \pi^*(\theta, \sigma^2|y) d\sigma^2}.$$

Clearly, when the posterior is proper, these conditionals are the usual ones based on $\pi(\theta, \sigma^2|y)$. On the other hand, when the posterior is improper, they are incompatible conditional densities; i.e., there is no (proper) joint density that generates them. This incompatibility does not prevent us from using these conditionals to run the Gibbs sampler as described in the Introduction, but the resulting Markov chain will be unstable. This will be formalized below.

We now describe a minimal set of conditions under which the integrability conditions are satisfied. Let $\hat{\theta} = (W^T W)^{-1} W^T y$, and assume that

$$2b_e + \|y - W\hat{\theta}\|^2 > 0. \quad (4)$$

Note that $\|Y - W\hat{\theta}\|^2$ is strictly positive with probability one under the data generating model. Also, (4) implies that, for all $\theta \in \mathbb{R}^{p+q}$,

$$2b_e + \|y - W\theta\|^2 = 2b_e + \|y - W\hat{\theta}\|^2 + \|W\theta - W\hat{\theta}\|^2 \geq 2b_e + \|y - W\hat{\theta}\|^2 > 0.$$

Define $\tilde{s} = \min \{q_1 + 2a_1, q_2 + 2a_2, \dots, q_r + 2a_r, N + 2a_e\}$, and assume that

$$\tilde{s} > 0. \quad (5)$$

Finally, assume that

$$b_i \geq 0 \quad \forall i \in \{1, 2, \dots, r\}. \quad (6)$$

Under (4), (5) and (6), the integrability conditions, (2) and (3), hold, so the conditional densities are well defined. Routine manipulation of $\pi^*(\theta, \sigma^2|y)$ shows that $\pi(\theta|\sigma^2, y)$ is a multivariate normal density with mean vector

$$m = \begin{bmatrix} (X^T X)^{-1} X^T (I - (\sigma_e^2)^{-1} Z Q^{-1} Z^T P^\perp) y \\ (\sigma_e^2)^{-1} Q^{-1} Z^T P^\perp y \end{bmatrix},$$

and covariance matrix

$$V = \begin{bmatrix} ((\sigma_e^2)^{-1} X^T X)^{-1} + (X^T X)^{-1} X^T Z Q^{-1} Z^T X (X^T X)^{-1} & -(X^T X)^{-1} X^T Z Q^{-1} \\ -Q^{-1} Z^T X (X^T X)^{-1} & Q^{-1} \end{bmatrix},$$

where $P^\perp = I - P = I - X(X^T X)^{-1} X^T$ and $Q = (\sigma_e^2)^{-1} Z^T P^\perp Z + D^{-1}$.

Things are a bit more complicated for $\pi(\sigma^2|\theta, y)$ due to the possible existence of a bothersome set of measure zero. Define $A = \{i \in \{1, 2, \dots, r\} : b_i = 0\}$. If A is empty, then $\pi(\sigma^2|\theta, y)$ is well defined for every $\theta \in \mathbb{R}^{p+q}$, and it is the following product of $r + 1$ inverted gamma densities

$$\pi(\sigma^2|\theta, y) = f_{\text{IG}}\left(\sigma_e^2; \frac{N}{2} + a_e, b_e + \frac{\|y - W\theta\|^2}{2}\right) \prod_{i=1}^r f_{\text{IG}}\left(\sigma_{u_i}^2; \frac{q_i}{2} + a_i, b_i + \frac{\|u_i\|^2}{2}\right),$$

where

$$f_{\text{IG}}(v; c, d) = \begin{cases} \frac{d^c}{\Gamma(c) v^{c+1}} e^{-d/v} & v > 0 \\ 0 & v \leq 0, \end{cases}$$

for $c, d > 0$. On the other hand, if A is nonempty, then

$$\int_{\mathbb{R}_+^{r+1}} \pi^*(\theta, \sigma^2|y) d\sigma^2 = \infty$$

whenever $\theta \in \mathcal{N} := \{\theta \in \mathbb{R}^{p+q} : \prod_{i \in A} \|u_i\| = 0\}$. The fact that $\pi(\sigma^2|\theta, y)$ is not defined when $\theta \in \mathcal{N}$ is irrelevant from a simulation standpoint because the probability of observing a θ in \mathcal{N} is zero. However, in order to perform a theoretical analysis, the Markov transition density (Mtd) of the Gibbs Markov chain must be defined for *every* $\theta \in \mathbb{R}^{p+q}$. Obviously, the Mtd can be defined arbitrarily on a set of measure zero. We define it as follows:

$$\pi(\sigma^2|\theta, y) = \begin{cases} f_{\text{IG}}\left(\sigma_e^2; \frac{N}{2} + a_e, b_e + \frac{\|y - W\theta\|^2}{2}\right) \prod_{i=1}^r f_{\text{IG}}\left(\sigma_{u_i}^2; \frac{q_i}{2} + a_i, b_i + \frac{\|u_i\|^2}{2}\right) & \text{if } \theta \notin \mathcal{N} \\ f_{\text{IG}}(\sigma_e^2; 1, 1) \prod_{i=1}^r f_{\text{IG}}(\sigma_{u_i}^2; 1, 1) & \text{if } \theta \in \mathcal{N}. \end{cases}$$

This definition can be used in all cases if we simply define \mathcal{N} to be \emptyset whenever A is empty.

The Mtd of the Gibbs Markov chain, $\{(\theta_n, \sigma_n^2)\}_{n=0}^\infty$, is defined as

$$k(\theta, \sigma^2|\tilde{\theta}, \tilde{\sigma}^2) = \pi(\sigma^2|\theta, y) \pi(\theta|\tilde{\sigma}^2, y).$$

It's easy to see that the chain is ψ -irreducible, and that $\pi^*(\theta, \sigma^2|y)$ is an unnormalized invariant density. It follows that the chain is positive recurrent if and only if the posterior is proper (Meyn and Tweedie, 1993, Chapter 10). Since a geometrically ergodic chain is necessarily positive recurrent, the Gibbs Markov chain cannot be geometrically ergodic when the posterior is improper.

The marginal sequences, $\{\theta_n\}_{n=0}^\infty$ and $\{\sigma_n^2\}_{n=0}^\infty$, are themselves Markov chains (see, e.g., Liu et al., 1994). The σ^2 -chain lives on \mathbb{R}_+^{r+1} and has Mtd given by

$$k_1(\sigma^2|\tilde{\sigma}^2) = \int_{\mathbb{R}^{p+q}} \pi(\sigma^2|\theta, y) \pi(\theta|\tilde{\sigma}^2, y) d\theta,$$

and invariant density $\int_{\mathbb{R}^{p+q}} \pi^*(\theta, \sigma^2|y) d\theta$. Similarly, the θ -chain lives on \mathbb{R}^{p+q} and has Mtd

$$k_2(\theta|\tilde{\theta}) = \int_{\mathbb{R}_+^{r+1}} \pi(\theta|\sigma^2, y) \pi(\sigma^2|\tilde{\theta}, y) d\sigma^2,$$

and invariant density $\int_{\mathbb{R}_+^{r+1}} \pi^*(\theta, \sigma^2|y) d\sigma^2$. Since the two marginal chains are also ψ -irreducible, they are positive recurrent if and only if the posterior is proper. Moreover, when the posterior is proper, routine calculations show that all three chains are Harris ergodic; i.e., positive Harris recurrent, ψ -irreducible and aperiodic (see Román (2012) for details). An important fact that we will exploit is that geometric ergodicity is a solidarity property for the three chains $\{(\theta_n, \sigma_n^2)\}_{n=0}^\infty$, $\{\theta_n\}_{n=0}^\infty$ and $\{\sigma_n^2\}_{n=0}^\infty$; that is, either all three are geometric or none of them is (Diaconis et al., 2008; Roberts and Rosenthal, 2001). In the next section, we prove that the Gibbs Markov chain converges at a geometric rate by proving that one of the marginal chains does.

3 The Main Result

In order to state the main result, we need a bit more notation. Write the spectral decomposition of the non-negative definite matrix $Z^T P^\perp Z$ as $\Gamma^T \Lambda \Gamma$, so Γ is a q -dimensional orthogonal matrix, and Λ is a diagonal matrix containing the eigenvalues of $Z^T P^\perp Z$, which we denote by $\{\lambda_i\}_{i=1}^q$. Define H to be a $q \times q$ diagonal matrix whose diagonal elements, $\{h_i\}_{i=1}^q$, are given by

$$h_i = \begin{cases} 1 & \lambda_i = 0 \\ 0 & \lambda_i \neq 0. \end{cases}$$

Finally, for $i \in \{1, \dots, r\}$, define R_i to be the $q_i \times q$ matrix of 0s and 1's such that $R_i u = u_i$. In other words, R_i is the matrix that *extracts* u_i from u . Here is our main result.

Theorem 2. *Assume that $\text{rank}(X) = p$. Assume further that $2b_e + \|y - W\hat{\theta}\|^2 > 0$, $\tilde{s} > 0$ and $b_i \geq 0$ for each $i \in \{1, 2, \dots, r\}$, so that the Gibbs sampler is well defined. If the following two conditions hold, then the Gibbs Markov chain is geometrically ergodic.*

1. *For each $i \in \{1, 2, \dots, r\}$, one of the following holds:*

$$(i) \ a_i < b_i = 0; \quad (ii) \ b_i > 0.$$

2. *There exists an $s \in (0, 1] \cap (0, \tilde{s}/2)$ such that*

$$2^{-s}(p+t)^s \frac{\Gamma(\frac{N}{2} + a_e - s)}{\Gamma(\frac{N}{2} + a_e)} < 1,$$

and

$$2^{-s} \sum_{i=1}^r \left\{ \frac{\Gamma(\frac{q_i}{2} + a_i - s)}{\Gamma(\frac{q_i}{2} + a_i)} \right\} \left(\text{tr}(R_i \Gamma^T H \Gamma R_i^T) \right)^s < 1 ,$$

where $t = \text{rank}(Z^T P^\perp Z)$.

Remark 1. It is important to reiterate that, by themselves, (4), (5) and (6) do not imply that the posterior density is proper. Of course, if conditions 1. and 2. in Theorem 2 hold as well, then the chain is geometric, so the posterior is necessarily proper.

Remark 2. A numerical search could be employed to check the second condition of Theorem 2. Indeed, one could evaluate the two functions of s at all points on a fine grid of values in the interval $(0, 1] \cap (0, \tilde{s}/2)$. The goal, of course, would be to find a single value of s at which both functions take values less than 1. Also, recall from the Introduction that Corollary 1 provides an alternative set of sufficient conditions that are harder to satisfy, but easier to check. A proof of Corollary 1 is given at the end of this section.

We will prove Theorem 2 indirectly by proving that the σ^2 -chain is geometrically ergodic (when the conditions of Theorem 2 hold). This is accomplished by establishing a *geometric drift condition* for the σ^2 -chain.

Proposition 1. Assume that $\text{rank}(X) = p$. Assume further that $2b_e + \|y - W\hat{\theta}\|^2 > 0$, $\tilde{s} > 0$ and $b_i \geq 0$ for each $i \in \{1, 2, \dots, r\}$. Under the two conditions of Theorem 2, there exist a $\rho \in [0, 1)$ and a finite constant L such that, for every $\tilde{\sigma}^2 \in \mathbb{R}_+^{r+1}$,

$$E(v(\sigma^2)|\tilde{\sigma}^2) \leq \rho v(\tilde{\sigma}^2) + L , \quad (7)$$

where the drift function is defined as

$$v(\sigma^2) = \alpha(\sigma_e^2)^s + \sum_{i=1}^r (\sigma_{u_i}^2)^s + \alpha(\sigma_e^2)^{-c} + \sum_{i=1}^r (\sigma_{u_i}^2)^{-c} ,$$

and $\alpha > 0$ and $c > 0$ are fixed constants to be determined. Hence, under the two conditions of Theorem 2, the σ^2 -chain is geometrically ergodic.

Proof. By conditioning on θ and iterating, we can express $E(v(\sigma^2)|\tilde{\sigma}^2)$ as

$$E \left[\alpha E((\sigma_e^2)^s | \theta) + E \left(\sum_{i=1}^r (\sigma_{u_i}^2)^s | \theta \right) + \alpha E((\sigma_e^2)^{-c} | \theta) + E \left(\sum_{i=1}^r (\sigma_{u_i}^2)^{-c} | \theta \right) \middle| \tilde{\sigma}^2 \right] . \quad (8)$$

We now develop upper bounds for each of the four terms inside the square brackets in (8). Fix $s \in S := (0, 1] \cap (0, \tilde{s}/2)$, and define

$$G_0(s) = 2^{-s} \frac{\Gamma(\frac{N}{2} + a_e - s)}{\Gamma(\frac{N}{2} + a_e)},$$

and, for each $i \in \{1, 2, \dots, r\}$, define

$$G_i(s) = 2^{-s} \frac{\Gamma(\frac{q_i}{2} + a_i - s)}{\Gamma(\frac{q_i}{2} + a_i)}.$$

Note that, since $s \in (0, 1]$, $(x_1 + x_2)^s \leq x_1^s + x_2^s$ whenever $x_1, x_2 \geq 0$. Thus,

$$\begin{aligned} E((\sigma_e^2)^s | \theta) &= 2^s G_0(s) \left(b_e + \frac{\|y - W\theta\|^2}{2} \right)^s \\ &\leq 2^s G_0(s) \left[b_e^s + \left(\frac{\|y - W\theta\|^2}{2} \right)^s \right] \\ &= G_0(s) (\|y - W\theta\|^2)^s + \text{const}, \end{aligned}$$

where “const” denotes a generic constant. Similarly,

$$E((\sigma_{u_i}^2)^s | \theta) = 2^s G_i(s) \left(b_i + \frac{\|u_i\|^2}{2} \right)^s \leq G_i(s) (\|u_i\|^2)^s + \text{const}.$$

Now, for any $c > 0$, we have

$$\begin{aligned} E((\sigma_e^2)^{-c} | \theta) &= 2^{-c} G_0(-c) \left(b_e + \frac{\|y - W\theta\|^2}{2} \right)^{-c} \\ &\leq 2^{-c} G_0(-c) \left(b_e + \frac{\|y - W\hat{\theta}\|^2}{2} \right)^{-c} = \text{const}, \end{aligned}$$

and, for each $i \in \{1, 2, \dots, r\}$,

$$\begin{aligned} E((\sigma_{u_i}^2)^{-c} | \theta) &= 2^{-c} G_i(-c) \left(b_i + \frac{\|u_i\|^2}{2} \right)^{-c} \\ &\leq G_i(-c) \left[(\|u_i\|^2)^{-c} I_{\{0\}}(b_i) + (2b_i)^{-c} I_{(0, \infty)}(b_i) \right]. \end{aligned}$$

Let $A = \{i : a_i < b_i = 0\}$, and note that $E\left(\sum_{i=1}^r (\sigma_{u_i}^2)^{-c} | \theta\right)$ can be bounded above by a constant if A is empty. Thus, we consider the case in which A is empty separately from the case where $A \neq \emptyset$. We begin with the latter, which is the more difficult case.

Case I: A is non-empty. Combining the four bounds above (and applying Jensen’s inequality twice), we have

$$\begin{aligned} E(v(\sigma^2) | \tilde{\sigma}^2) &\leq \alpha G_0(s) \left[E(\|y - W\theta\|^2 | \tilde{\sigma}^2) \right]^s + \sum_{i=1}^r G_i(s) \left[E(\|u_i\|^2 | \tilde{\sigma}^2) \right]^s \\ &\quad + \sum_{i \in A} G_i(-c) E\left[\|u_i\|^{-2c} | \tilde{\sigma}^2\right] + \text{const}. \end{aligned} \tag{9}$$

Appendix A.2 contains a proof of the following inequality:

$$E[\|y - W\theta\|^2 | \tilde{\sigma}^2] \leq (p + t) \tilde{\sigma}_e^2 + \text{const} . \quad (10)$$

It follows immediately that

$$\left[E(\|y - W\theta\|^2 | \tilde{\sigma}^2) \right]^s \leq (p + t)^s (\tilde{\sigma}_e^2)^s + \text{const} .$$

In Appendix A.3, it is shown that, for each $i \in \{1, 2, \dots, r\}$, we have

$$E[\|u_i\|^2 | \tilde{\sigma}^2] \leq \xi_i \tilde{\sigma}_e^2 + \zeta_i \sum_{j=1}^r \tilde{\sigma}_{u_j}^2 + \text{const} ,$$

where $\xi_i = \text{tr}(R_i(Z^T P^\perp Z)^+ R_i^T)$, $\zeta_i = \text{tr}(R_i \Gamma^T H \Gamma R_i^T)$, and A^+ denotes the Moore-Penrose inverse of the matrix A . It follows that

$$\left[E(\|u_i\|^2 | \tilde{\sigma}^2) \right]^s \leq \xi_i^s (\tilde{\sigma}_e^2)^s + \zeta_i^s \sum_{j=1}^r (\tilde{\sigma}_{u_j}^2)^s + \text{const} . \quad (11)$$

In Appendix A.4, it is established that, for any $c \in (0, 1/2)$, and for each $i \in \{1, 2, \dots, r\}$, we have

$$E[\|u_i\|^{-2c} | \tilde{\sigma}^2] \leq 2^{-c} \frac{\Gamma(\frac{q_i}{2} - c)}{\Gamma(\frac{q_i}{2})} \left[\lambda_{\max}^c (\tilde{\sigma}_e^2)^{-c} + (\tilde{\sigma}_{u_i}^2)^{-c} \right] , \quad (12)$$

where λ_{\max} denotes the largest eigenvalue of $Z^T P^\perp Z$. Combining (10) - (12) with (9), we have

$$\begin{aligned} E(v(\sigma^2) | \tilde{\sigma}^2) &\leq \alpha \left(\delta_1(s) + \frac{\delta_2(s)}{\alpha} \right) (\tilde{\sigma}_e^2)^s + \delta_3(s) \sum_{j=1}^r (\tilde{\sigma}_{u_j}^2)^s \\ &\quad + \alpha \frac{\delta_4(c)}{\alpha} (\tilde{\sigma}_e^2)^{-c} + \delta_5(c) \sum_{j \in A} (\tilde{\sigma}_{u_j}^2)^{-c} + \text{const} , \end{aligned} \quad (13)$$

where

$$\begin{aligned} \delta_1(s) &:= G_0(s)(p + t)^s , \quad \delta_2(s) := \sum_{i=1}^r \xi_i^s G_i(s) , \quad \delta_3(s) := \sum_{i=1}^r \zeta_i^s G_i(s) , \\ \delta_4(c) &:= 2^{-c} \lambda_{\max}^c \sum_{i \in A} G_i(-c) \frac{\Gamma(\frac{q_i}{2} - c)}{\Gamma(\frac{q_i}{2})} \quad \text{and} \quad \delta_5(c) := 2^{-c} \max_{i \in A} \left[G_i(-c) \frac{\Gamma(\frac{q_i}{2} - c)}{\Gamma(\frac{q_i}{2})} \right] . \end{aligned}$$

Hence,

$$E(v(\sigma^2) | \tilde{\sigma}^2) \leq \rho(\alpha, s, c) v(\tilde{\sigma}^2) + L ,$$

where

$$\rho(\alpha, s, c) = \max \left\{ \delta_1(s) + \frac{\delta_2(s)}{\alpha}, \delta_3(s), \frac{\delta_4(c)}{\alpha}, \delta_5(c) \right\} .$$

All that remains is to show that there exists a triple $(\alpha, s, c) \in \mathbb{R}_+ \times S \times (0, 1/2)$ such that $\rho(\alpha, s, c) < 1$. We begin by demonstrating that, if c is small enough, then $\delta_5(c) < 1$. Define $\tilde{a} = -\max_{i \in A} a_i$. Also, set $C = (0, 1/2) \cap (0, \tilde{a})$. Fix $c \in C$ and note that

$$\delta_5(c) = \max_{i \in A} \left[\frac{\Gamma(\frac{q_i}{2} + a_i + c)}{\Gamma(\frac{q_i}{2} + a_i)} \frac{\Gamma(\frac{q_i}{2} - c)}{\Gamma(\frac{q_i}{2})} \right].$$

For any $i \in A$, $c + a_i < 0$, and since $\tilde{s} > 0$, it follows that

$$0 < \frac{q_i}{2} + a_i < \frac{q_i}{2} + a_i + c < \frac{q_i}{2}.$$

But, $\Gamma(x - z)/\Gamma(x)$ is decreasing in x for $x > z > 0$, so we have,

$$\frac{\Gamma(\frac{q_i}{2} + a_i)}{\Gamma(\frac{q_i}{2} + a_i + c)} = \frac{\Gamma(\frac{q_i}{2} + a_i + c - c)}{\Gamma(\frac{q_i}{2} + a_i + c)} > \frac{\Gamma(\frac{q_i}{2} - c)}{\Gamma(\frac{q_i}{2})},$$

and it follows immediately that $\delta_5(c) < 1$ whenever $c \in C$. The two conditions of Theorem 2 imply that there exists an $s^* \in S$ such that $\delta_1(s^*) < 1$ and $\delta_3(s^*) < 1$. Let c^* be any point in C , and choose α^* to be any number larger than

$$\max \left\{ \frac{\delta_2(s^*)}{1 - \delta_1(s^*)}, \delta_4(c^*) \right\}.$$

A simple calculation shows that $\rho(\alpha^*, s^*, c^*) < 1$, and this completes the proof for Case I.

Case II: $A = \emptyset$. Since we no longer have to deal with $E((\sigma_{u_i}^2)^{-c} | \theta)$, the bound (13) becomes

$$E(v(\sigma^2) | \tilde{\sigma}^2) \leq \alpha \left(\delta_1(s) + \frac{\delta_2(s)}{\alpha} \right) (\tilde{\sigma}_e^2)^s + \delta_3(s) \sum_{j=1}^r (\tilde{\sigma}_{u_j}^2)^s + \text{const},$$

and there is no restriction on c other than $c > 0$. Hence,

$$E(v(\sigma^2) | \tilde{\sigma}^2) \leq \rho(\alpha, s) v(\tilde{\sigma}^2) + L,$$

where

$$\rho(\alpha, s) = \max \left\{ \delta_1(s) + \frac{\delta_2(s)}{\alpha}, \delta_3(s) \right\}.$$

All that remains is to show that there exists a $(\alpha, s) \in \mathbb{R}_+ \times S$ such that $\rho(\alpha, s) < 1$. As in Case I, the two conditions of Proposition 2 imply that there exists an $s^* \in S$ such that $\delta_1(s^*) < 1$ and $\delta_3(s^*) < 1$. Let α^* be any number larger than

$$\frac{\delta_2(s^*)}{1 - \delta_1(s^*)}.$$

A simple calculation shows that $\rho(\alpha^*, s^*) < 1$, and this completes the proof for Case II.

The only thing left to do is to show that (7) implies that the σ^2 -chain is geometrically ergodic. Note that the σ^2 -chain is ψ -irreducible and aperiodic. Moreover, because its Mtd is strictly positive on $\mathbb{R}_+^{r+1} \times \mathbb{R}_+^{r+1}$, its maximal irreducibility measure is equivalent to Lebesgue measure on \mathbb{R}_+^{r+1} . Thus, it follows from Meyn and Tweedie's (1993) Lemma 15.2.8 and their Theorem 6.0.1 that, if the σ^2 -chain is a Feller chain and the drift function is unbounded off compact sets, then (7) implies that the σ^2 -chain is geometrically ergodic.

We first show that the drift function is unbounded off compact sets; i.e., we will demonstrate that, for every $d \in \mathbb{R}$, the set

$$S_d = \{\sigma^2 \in \mathbb{R}_+^{r+1} : v(\sigma^2) \leq d\}$$

is compact. Let d be such that S_d is non-empty (otherwise S_d is trivially compact), which means that d and d/α must be larger than 1. Since $v(\sigma^2)$ is a continuous function, S_d is closed in \mathbb{R}_+^{r+1} . Now consider the following set:

$$T_d = [(d/\alpha)^{-1/c}, (d/\alpha)^{1/s}] \times [d^{-1/c}, d^{1/s}] \times \dots \times [d^{-1/c}, d^{1/s}] .$$

The set T_d is compact in \mathbb{R}_+^{r+1} , and hence in \mathbb{R}_+^{r+1} . Since $S_d \subset T_d$, S_d is a closed subset of a compact set in \mathbb{R}_+^{r+1} , so it is compact in \mathbb{R}_+^{r+1} . Hence, the drift function is unbounded off compact sets.

To complete the argument, we must show that the σ^2 -chain is a Feller chain. Let P_1 denote the Markov transition function of the σ^2 -chain; that is, for any $\tilde{\sigma}^2 \in \mathbb{R}_+^{r+1}$ and any Borel set A ,

$$P_1(\tilde{\sigma}^2, A) = \int_A k_1(\sigma^2 | \tilde{\sigma}^2) d\sigma^2 .$$

The chain is Feller if, for each fixed open set O , $P_1(\cdot, O)$ is a lower semi-continuous function on \mathbb{R}_+^{r+1} . To this end, let $\{\tilde{\sigma}_m^2\}_{m=1}^\infty$ be a sequence in \mathbb{R}_+^{r+1} that converges to $\tilde{\sigma}^2 \in \mathbb{R}_+^{r+1}$. Then

$$\begin{aligned} \liminf_{m \rightarrow \infty} P_1(\tilde{\sigma}_m^2, O) &= \liminf_{m \rightarrow \infty} \int_O k_1(\sigma^2 | \tilde{\sigma}_m^2) d\sigma^2 \\ &= \liminf_{m \rightarrow \infty} \int_O \left[\int_{\mathbb{R}^{p+q}} \pi(\sigma^2 | \theta, y) \pi(\theta | \tilde{\sigma}_m^2, y) d\theta \right] d\sigma^2 \\ &\geq \int_O \int_{\mathbb{R}^{p+q}} \pi(\sigma^2 | \theta, y) \left[\liminf_{m \rightarrow \infty} \pi(\theta | \tilde{\sigma}_m^2, y) \right] d\theta d\sigma^2 \\ &= \int_O \left[\int_{\mathbb{R}^{p+q}} \pi(\sigma^2 | \theta, y) \pi(\theta | \tilde{\sigma}^2, y) d\theta \right] d\sigma^2 \\ &= P_1(\tilde{\sigma}^2, O) , \end{aligned}$$

where the inequality follows from Fatou's Lemma, and the third equality follows from the fact that $\pi(\theta|\sigma^2, y)$ is continuous in σ^2 . (For a proof of continuity, see Román (2012).) We conclude that $P_1(\cdot, O)$ is lower semi-continuous, so the σ^2 -chain is Feller. The proof is now complete. \square

We end this section with a proof of Corollary 1.

Proof of Corollary 1. It suffices to show that, together, conditions (B') and (C') of Corollary 1 imply the second condition of Theorem 2. Clearly, (B') and (C') imply that $\tilde{s}/2 > 1$, so $(0, 1] \cap (0, \tilde{s}/2) = (0, 1]$. Take $s^* = 1$. Condition (C') implies

$$2^{-s^*} (p+t)^{s^*} \frac{\Gamma(\frac{N}{2} + a_e - s^*)}{\Gamma(\frac{N}{2} + a_e)} = \frac{p+t}{N+2a_e-2} < 1.$$

Now, since $\text{tr}(H)$ is the number of zero eigenvalues of $Z^T P^\perp Z$, we have

$$\sum_{i=1}^r \text{tr}(R_i \Gamma^T H \Gamma R_i^T) = \text{tr}\left(\Gamma^T H \Gamma \sum_{i=1}^r R_i^T R_i\right) = \text{tr}(\Gamma^T H \Gamma I) = \text{tr}(H) = q - t.$$

Hence,

$$\begin{aligned} 2^{-s^*} \sum_{i=1}^r \left\{ \frac{\Gamma(\frac{q_i}{2} + a_i - s^*)}{\Gamma(\frac{q_i}{2} + a_i)} \right\} \left(\text{tr}(R_i \Gamma^T H \Gamma R_i^T) \right)^{s^*} &= \sum_{i=1}^r \frac{\text{tr}(R_i \Gamma^T H \Gamma R_i^T)}{q_i + 2a_i - 2} \\ &\leq \frac{\sum_{i=1}^r \text{tr}(R_i \Gamma^T H \Gamma R_i^T)}{\min_{j \in \{1, 2, \dots, r\}} \{q_j + 2a_j - 2\}} \\ &= \frac{q - t}{\min_{j \in \{1, 2, \dots, r\}} \{q_j + 2a_j - 2\}} \\ &< 1, \end{aligned}$$

where the last inequality follows from condition (B') . \square

4 Discussion

Our decision to work with the σ^2 -chain rather than the θ -chain was based on an important technical difference between the two chains that stems from the fact that $\pi(\sigma^2|\theta, y)$ is not continuous in θ for each fixed σ^2 . Indeed, recall that

$$\pi(\sigma^2|\theta, y) = \begin{cases} f_{\text{IG}}\left(\sigma_e^2; \frac{N}{2} + a_e, b_e + \frac{\|y - W\theta\|^2}{2}\right) \prod_{i=1}^r f_{\text{IG}}\left(\sigma_{u_i}^2; \frac{q_i}{2} + a_i, b_i + \frac{\|u_i\|^2}{2}\right) & \text{if } \theta \notin \mathcal{N} \\ f_{\text{IG}}(\sigma_e^2; 1, 1) \prod_{i=1}^r f_{\text{IG}}(\sigma_{u_i}^2; 1, 1) & \text{if } \theta \in \mathcal{N}. \end{cases}$$

Also, recall that the Mtd of the σ^2 -chain is given by

$$k_1(\sigma^2|\tilde{\sigma}^2) = \int_{\mathbb{R}^{p+q}} \pi(\sigma^2|\theta, y) \pi(\theta|\tilde{\sigma}^2, y) d\theta .$$

Since the set \mathcal{N} has measure zero, the “arbitrary part” of $\pi(\sigma^2|\theta, y)$ washes out of k_1 . However, the same cannot be said for the θ -chain, whose Mtd is given by

$$k_2(\theta|\tilde{\theta}) = \int_{\mathbb{R}_+^{r+1}} \pi(\theta|\sigma^2, y) \pi(\sigma^2|\tilde{\theta}, y) d\sigma^2 .$$

This difference between k_1 and k_2 comes into play when we attempt to apply certain “topological” results from Markov chain theory, such as those in Chapter 6 of Meyn and Tweedie (1993). In particular, in our proof that the σ^2 -chain is a Feller chain (which was part of the proof of Proposition 1), we used the fact that $\pi(\theta|\sigma^2, y)$ is continuous in σ^2 for each fixed θ . Since $\pi(\sigma^2|\theta, y)$ is not continuous, we cannot use the same argument to prove that the θ -chain is Feller. In fact, we suspect that the θ -chain is not Feller, and if this is true, it means that our method of establishing the applicability of Meyn and Tweedie’s (1993) Lemma 15.2.8 will not work for the θ -chain.

It is possible to circumvent the problem described above by removing the set \mathcal{N} from the state space of the θ -chain. In this case, we are no longer required to define $\pi(\sigma^2|\theta, y)$ for $\theta \in \mathcal{N}$, and since $\pi(\sigma^2|\theta, y)$ is continuous (for fixed σ^2) on $\mathbb{R}^{p+q} \setminus \mathcal{N}$, the Feller argument for the θ -chain will go through. On the other hand, the new state space has “holes” in it, and this could complicate the search for a drift function that is unbounded off compact sets. For example, consider a toy drift function given by $v(x) = x^2$. This function is clearly unbounded off compact sets when the state space is \mathbb{R} , but not when the state space is $\mathbb{R} \setminus \{0\}$. The modified drift function $v^*(x) = x^2 + 1/x^2$ is unbounded off compact sets for the “holey” state space.

Tan and Hobert (2009) overlooked a set of measure zero (similar to our \mathcal{N}), and this oversight led to an error in the proof of their Proposition 3. As we now explain, their proof can be repaired using the ideas described in the previous paragraph. Our work shows that their Proposition 3 is actually correct as stated. Recall that Tan and Hobert (2009) (hereafter, T&H) considered the centered version of the one-way random effects model, which, in their notation, is

$$Y_{ij} = \theta_i + \epsilon_{ij} ,$$

where $i = 1, \dots, q$, $j = 1, \dots, m_i$, the θ_i s are iid $N(\mu, \sigma_\theta^2)$ and the ϵ_{ij} s, which are independent of the θ_i s, are iid $N(0, \sigma_e^2)$. They considered a parametric family of improper prior densities given by

$$\pi_{a,b}(\mu, \sigma_\theta^2, \sigma_e^2) = (\sigma_\theta^2)^{-(a+1)} (\sigma_e^2)^{-(b+1)} ,$$

where a and b are known hyper-parameters. Let $\sigma^2 = (\sigma_\theta^2, \sigma_e^2)$ and $\xi = (\mu, \theta_1, \dots, \theta_q)$. T&H analyzed the ξ -chain, $\{\xi_n\}_{n=0}^\infty$, which was defined to have state space \mathbb{R}^{q+1} , and Mtd

$$\hat{k}(\tilde{\xi}|\xi) = \int_{\mathbb{R}_+^2} \pi(\tilde{\xi}|\sigma^2, y) \pi(\sigma^2|\xi, y) d\sigma^2 .$$

The conditional density $\pi(\sigma^2|\xi, y)$ is the product of two inverted gamma densities, one of which is the source of the problem. Indeed, $\pi(\sigma_\theta^2|\xi, y)$ is inverted gamma with shape $q/2 + a$ and scale

$$\frac{1}{2} \sum_{i=1}^q (\theta_i - \mu)^2 .$$

T&H overlooked the fact that this density is not defined on the set

$$\mathcal{N}^* = \{\xi \in \mathbb{R}^{q+1} : \mu = \theta_1 = \dots = \theta_q\} .$$

Thus, the Mtd \hat{k} is not well-defined for $\xi \in \mathcal{N}^*$, and, as a result, T&H's argument showing that the ξ -chain is Feller (as a chain on \mathbb{R}^{q+1}) is incorrect.

T&H's proof can be repaired by redefining the state space of the ξ chain to be $\mathbb{R}^{q+1} \setminus \mathcal{N}^*$. For fixed σ^2 , $\pi(\sigma^2|\xi, y)$ is a continuous function of ξ on $\mathbb{R}^{q+1} \setminus \mathcal{N}^*$, and it follows that the ξ -chain is Feller on the new state space. Now, the drift function that is used in T&H's proof of Proposition 3 takes the form

$$\epsilon \left[\sum_{i=1}^q (\theta_i - \mu)^2 \right]^s + \left[\sum_{i=1}^q m_i (\bar{y}_i - \theta_i)^2 \right]^s ,$$

where $\epsilon > 0$ and $s \in (0, 1]$. This function is unbounded off compact sets when the state space is \mathbb{R}^{q+1} , but not when the state space is $\mathbb{R}^{q+1} \setminus \mathcal{N}^*$. To remedy this problem, we add the following term to the drift function

$$\left[\sum_{i=1}^q (\theta_i - \mu)^2 \right]^{-c} ,$$

where $c > 0$. Since this function blows up as ξ approaches \mathcal{N}^* , the modified drift function is unbounded off compact sets on the new state space, $\mathbb{R}^{q+1} \setminus \mathcal{N}^*$. Straightforward calculations (using techniques similar to those employed in our Appendix A.4) show that the modified drift function still satisfies a geometric drift condition, which implies that the ξ -chain defined on $\mathbb{R}^{q+1} \setminus \mathcal{N}^*$ is geometrically ergodic. Moreover, this result, in conjunction with Meyn and Tweedie's (1993) Theorem 15.0.1, implies that the original ξ -chain (defined on \mathbb{R}^{q+1}) is also geometric. Thus, T&H's

result is correct as stated. For a more detailed version of the corrected proof, as well as an alternate proof based on the σ^2 -chain, see Román (2012).

Acknowledgment. The second author's work was supported by NSF Grants DMS-08-05860 and & DMS-11-06395.

Appendices

A Upper Bounds

A.1 Preliminary results

Here is our first result.

Lemma 1. *The following inequalities hold for all $\sigma^2 \in \mathbb{R}_+^{r+1}$ and all $i \in \{1, 2, \dots, r\}$:*

1. $Q^{-1} \preceq (Z^T P^\perp Z)^+ \sigma_e^2 + \Gamma^T H \Gamma (\sum_{j=1}^r \sigma_{u_j}^2)$
2. $\text{tr}(P^\perp Z Q^{-1} Z^T P^\perp) \leq \text{rank}(Z^T P^\perp Z) \sigma_e^2$
3. $(R_i Q^{-1} R_i^T)^{-1} \preceq ((\sigma_e^2)^{-1} \lambda_{\max} + (\sigma_{u_i}^2)^{-1}) I_{q_i}.$

Proof. Recall from Section 3 that $\Gamma^T \Lambda \Gamma$ is the spectral decomposition of $Z^T P^\perp Z$, and that H is a binary diagonal matrix whose i th diagonal element is 1 if and only if the i th diagonal element of Λ is 0. Let $\sigma_\bullet^2 = \sum_{j=1}^r \sigma_{u_j}^2$. Since $(\sigma_\bullet^2)^{-1} I_q \preceq D^{-1}$, we have

$$(\sigma_e^2)^{-1} Z^T P^\perp Z + (\sigma_\bullet^2)^{-1} I_q \preceq (\sigma_e^2)^{-1} Z^T P^\perp Z + D^{-1},$$

and this yields

$$\begin{aligned} Q^{-1} &= \left((\sigma_e^2)^{-1} Z^T P^\perp Z + D^{-1} \right)^{-1} \\ &\preceq \left((\sigma_e^2)^{-1} Z^T P^\perp Z + (\sigma_\bullet^2)^{-1} I_q \right)^{-1} \\ &= \Gamma^T \left(\Lambda (\sigma_e^2)^{-1} + I_q (\sigma_\bullet^2)^{-1} \right)^{-1} \Gamma. \end{aligned} \tag{14}$$

Now let Λ^+ be a diagonal matrix whose diagonal elements, $\{\lambda_i^+\}_{i=1}^q$, are given by

$$\lambda_i^+ = \begin{cases} \lambda_i^{-1} & \lambda_i \neq 0 \\ 0 & \lambda_i = 0. \end{cases}$$

Note that, for each $i \in \{1, 2, \dots, r\}$, we have

$$\frac{1}{\lambda_i(\sigma_e^2)^{-1} + (\sigma_\bullet^2)^{-1}} \leq \lambda_i^+ \sigma_e^2 + I_{\{0\}}(\lambda_i) \sigma_\bullet^2.$$

This shows that

$$\left(\Lambda(\sigma_e^2)^{-1} + I_q(\sigma_\bullet^2)^{-1} \right)^{-1} \preceq \Lambda^+ \sigma_e^2 + H \sigma_\bullet^2.$$

Together with (14), this leads to

$$Q^{-1} \preceq \Gamma^T \left(\Lambda(\sigma_e^2)^{-1} + I_q(\sigma_\bullet^2)^{-1} \right)^{-1} \Gamma \preceq \Gamma^T (\Lambda^+ \sigma_e^2 + H \sigma_\bullet^2) \Gamma = (Z^T P^\perp Z)^+ \sigma_e^2 + \Gamma^T H \Gamma \sigma_\bullet^2,$$

which proves the first statement. Now let $\tilde{Z} = P^\perp Z$. Pre- and post-multiplying the first statement by \tilde{Z} and \tilde{Z}^T , respectively, and then taking traces yields

$$\text{tr}(\tilde{Z} Q^{-1} \tilde{Z}^T) \leq \text{tr}(\tilde{Z}(\tilde{Z}^T \tilde{Z})^+ \tilde{Z}^T) \sigma_e^2 + \text{tr}(\tilde{Z} \Gamma^T H \Gamma \tilde{Z}^T) \sigma_\bullet^2. \quad (15)$$

Since $(\tilde{Z}^T \tilde{Z})(\tilde{Z}^T \tilde{Z})^+$ is idempotent, we have

$$\text{tr}(\tilde{Z}(\tilde{Z}^T \tilde{Z})^+ \tilde{Z}^T) = \text{tr}(\tilde{Z}^T \tilde{Z}(\tilde{Z}^T \tilde{Z})^+) = \text{rank}(\tilde{Z}^T \tilde{Z}(\tilde{Z}^T \tilde{Z})^+) = \text{rank}(\tilde{Z}^T \tilde{Z}).$$

Furthermore,

$$\text{tr}(\tilde{Z} \Gamma^T H \Gamma \tilde{Z}^T) = \text{tr}(\Gamma^T H \Gamma Z^T P^\perp Z) = \text{tr}(\Gamma^T H \Gamma \Gamma^T \Lambda \Gamma) = \text{tr}(\Gamma^T H \Lambda \Gamma) = 0,$$

where the last line follows from the fact that $H \Lambda = 0$. It follows from (15) that

$$\text{tr}(P^\perp Z Q^{-1} Z^T P^\perp) \leq \text{rank}(Z^T P^\perp Z) \sigma_e^2,$$

and the second statement has been established. Recall from Section 3 that λ_{\max} is the largest eigenvalue of $Z^T P^\perp Z$, and that R_i is the $q_i \times q$ matrix of 0s and 1's such that $R_i u = u_i$. Now, fix $i \in \{1, 2, \dots, r\}$ and note that

$$Q = (\sigma_e^2)^{-1} Z^T P^\perp Z + D^{-1} \preceq (\sigma_e^2)^{-1} \lambda_{\max} I_q + D^{-1}.$$

It follows that

$$R_i((\sigma_e^2)^{-1} \lambda_{\max} I_q + D^{-1})^{-1} R_i^T \preceq R_i Q^{-1} R_i^T,$$

and since these two matrices are both positive definite, we have

$$\begin{aligned} \left(R_i Q^{-1} R_i^T \right)^{-1} &\preceq \left(R_i((\sigma_e^2)^{-1} \lambda_{\max} I_q + D^{-1})^{-1} R_i^T \right)^{-1} \\ &= ((\sigma_e^2)^{-1} \lambda_{\max} + (\sigma_{u_i}^2)^{-1}) I_{q_i}, \end{aligned}$$

and this proves that the third statement is true. \square

Lemma 2. The function $h(\sigma^2) := \|(\sigma_e^2)^{-1}Q^{-1}Z^TP^\perp y\|$ is bounded on \mathbb{R}_+^{r+1} .

The following result from Khare and Hobert (2011) will be used in the proof of Lemma 2 .

Lemma 3. Fix $n \in \{2, 3, \dots\}$ and $m \in \mathbb{N}$, and let t_1, \dots, t_n be vectors in \mathbb{R}^m . Then

$$C_{m,n}(t_1; t_2, \dots, t_n) := \sup_{c \in \mathbb{R}_+^n} t_1^T \left(t_1 t_1^T + \sum_{i=2}^n c_i t_i t_i^T + c_1 I \right)^{-2} t_1$$

is finite.

Proof of Lemma 2. Let \tilde{z}_i and y_i denote the i th column of $\tilde{Z}^T = (P^\perp Z)^T$ and the i th component of y , respectively. Then,

$$\begin{aligned} h(\sigma^2) &= \left\| (Z^T P^\perp Z + \sigma_e^2 D^{-1})^{-1} Z^T P^\perp y \right\| \\ &= \left\| \sum_{i=1}^n (\tilde{Z}^T \tilde{Z} + \sigma_e^2 D^{-1})^{-1} \tilde{z}_i y_i \right\| \\ &\leq \sum_{i=1}^N \left\| (\tilde{Z}^T \tilde{Z} + \sigma_e^2 D^{-1})^{-1} \tilde{z}_i y_i \right\| \\ &= \sum_{i=1}^N \left\| \left(\sum_{j=1}^N \tilde{z}_j \tilde{z}_j^T + \sigma_e^2 D^{-1} \right)^{-1} \tilde{z}_i y_i \right\| \\ &= \sum_{i=1}^N |y_i| \left\| \left(\tilde{z}_i \tilde{z}_i^T + \sum_{j \in \{1, 2, \dots, N\} \setminus \{i\}} \tilde{z}_j \tilde{z}_j^T + \sigma_e^2 D^{-1} \right)^{-1} \tilde{z}_i \right\|. \end{aligned}$$

Therefore, it is enough to show that for each $i \in \{1, 2, \dots, N\}$,

$$K_i(\sigma^2) := \left\| \left(\tilde{z}_i \tilde{z}_i^T + \sum_{j \in \{1, 2, \dots, N\} \setminus \{i\}} \tilde{z}_j \tilde{z}_j^T + \sigma_e^2 D^{-1} \right)^{-1} \tilde{z}_i \right\|$$

is bounded. Now,

$$\begin{aligned} K_i^2(\sigma^2) &= \tilde{z}_i^T \left(\tilde{z}_i \tilde{z}_i^T + \sum_{j \in \{1, 2, \dots, N\} \setminus \{i\}} \tilde{z}_j \tilde{z}_j^T + \sigma_e^2 D^{-1} \right)^{-2} \tilde{z}_i \\ &= \tilde{z}_i^T \left(\tilde{z}_i \tilde{z}_i^T + \sum_{j \in \{1, 2, \dots, N\} \setminus \{i\}} \tilde{z}_j \tilde{z}_j^T + \sigma_e^2 \left(D^{-1} - \frac{1}{\sigma_\bullet^2} I_q \right) + \frac{\sigma_e^2}{\sigma_\bullet^2} I_q \right)^{-2} \tilde{z}_i \\ &\leq \sup_{c \in \mathbb{R}_+^{N+q}} t_i^T \left(t_i t_i^T + \sum_{j \in \{1, 2, \dots, N\} \setminus \{i\}} c_j t_j t_j^T + \sum_{j=N+1}^{N+q} c_j t_j t_j^T + c_i I_q \right)^{-2} t_i \end{aligned}$$

where, for $j = 1, 2, \dots, N$, $t_j = \tilde{z}_j$, and for $j \in \{N+1, \dots, N+q\}$, the t_j are the standard orthonormal basis vectors in \mathbb{R}^q ; that is, t_{N+l} has a one in the l th position and zeros everywhere else. An application of Lemma 3 shows that $K_i^2(\sigma^2)$ is bounded. Hence, $h(\sigma^2)$ is bounded. \square

Let $\chi_k^2(w)$ denote the non-central chi-square distribution with k degrees of freedom and non-centrality parameter w .

Lemma 4. *If $U \sim \chi_k^2(w)$ and $\gamma \in (0, k/2)$, then*

$$E[U^{-\gamma}] \leq \frac{2^{-\gamma} \Gamma(\frac{k}{2} - \gamma)}{\Gamma(\frac{k}{2})}.$$

Proof. Since $\Gamma(x - \gamma)/\Gamma(x)$ is decreasing for $x > \gamma > 0$, we have

$$\begin{aligned} E[U^{-\gamma}] &= \sum_{i=0}^{\infty} \frac{w^i e^{-w}}{i!} \int_{\mathbb{R}_+} u^{-\gamma} \left[\frac{1}{\Gamma(\frac{k}{2} + i) 2^{\frac{k}{2} + i}} u^{\frac{k}{2} + i - 1} e^{-\frac{u}{2}} \right] du \\ &= 2^{-\gamma} \sum_{i=0}^{\infty} \frac{w^i e^{-w}}{i!} \frac{\Gamma(\frac{k}{2} + i - \gamma)}{\Gamma(\frac{k}{2} + i)} \\ &\leq 2^{-\gamma} \frac{\Gamma(\frac{k}{2} - \gamma)}{\Gamma(\frac{k}{2})}. \end{aligned}$$

□

A.2 An upper bound on $E[\|y - W\theta\|^2 | \sigma^2]$

Recall that $\theta = (\beta^T u^T)^T$, $W = (X \ Z)$, and that $\pi(\theta | \sigma^2, y)$ is a multivariate normal density with mean m and covariance matrix V . Thus,

$$E[\|y - W\theta\|^2 | \sigma^2] = \text{tr}(WVW^T) + \|y - Wm\|^2, \quad (16)$$

and we have,

$$\begin{aligned} \text{tr}(WVW^T) &= \sigma_e^2 \text{tr}(P) + \text{tr}(PZQ^{-1}Z^T P) - 2 \text{tr}(ZQ^{-1}Z^T P) + \text{tr}(ZQ^{-1}Z^T) \\ &= p\sigma_e^2 - \text{tr}(ZQ^{-1}Z^T P) + \text{tr}(ZQ^{-1}Z^T) \\ &= p\sigma_e^2 + \text{tr}(ZQ^{-1}Z^T P^\perp) \\ &= p\sigma_e^2 + \text{tr}(P^\perp ZQ^{-1}Z^T P^\perp) \\ &\leq p\sigma_e^2 + \text{rank}(Z^T P^\perp Z) \sigma_e^2 \\ &= (p + t) \sigma_e^2, \end{aligned} \quad (17)$$

where the inequality is an application of Lemma 1. Finally, a simple calculation shows that

$$y - Wm = P^\perp [I - (\sigma_e^2)^{-1} ZQ^{-1}Z^T P^\perp] y.$$

Hence,

$$\begin{aligned}
\|y - Wm\| &= \|P^\perp y - (\sigma_e^2)^{-1} P^\perp Z Q^{-1} Z^T P^\perp y\| \\
&\leq \|P^\perp y\| + \|(\sigma_e^2)^{-1} P^\perp Z Q^{-1} Z^T P^\perp y\| \\
&\leq \|P^\perp y\| + \|P^\perp Z\| \|(\sigma_e^2)^{-1} Q^{-1} Z^T P^\perp y\| \\
&\leq \text{const} ,
\end{aligned} \tag{18}$$

where $\|\cdot\|$ denotes the Frobenius norm and the last inequality uses Lemma 2. Finally, combining (16), (17) and (18) yields

$$E[\|y - W\theta\|^2 | \sigma^2] \leq (p + t) \sigma_e^2 + \text{const} .$$

A.3 An upper bound on $E[\|u_i\|^2 | \sigma^2]$

Note that

$$E[\|u_i\|^2 | \sigma^2] = E[\|R_i u\|^2 | \sigma^2] = \text{tr}(R_i Q^{-1} R_i^T) + \|E[R_i u | \sigma^2]\|^2 . \tag{19}$$

By Lemma 1, we have

$$\begin{aligned}
\text{tr}(R_i Q^{-1} R_i^T) &\leq \text{tr}(R_i (Z^T P^\perp Z)^+ R_i^T) \sigma_e^2 + \text{tr}(R_i \Gamma^T H \Gamma R_i^T) \sum_{j=1}^r \sigma_{u_j}^2 \\
&= \xi_i \sigma_e^2 + \zeta_i \sum_{j=1}^r \sigma_{u_j}^2 .
\end{aligned} \tag{20}$$

Now, by Lemma 2

$$\|E[R_i u | \sigma^2]\| \leq \|R_i\| \|E[u | \sigma^2]\| = \|R_i\| h(\sigma^2) \leq \text{const} . \tag{21}$$

Combining (19), (20) and (21) yields

$$E[\|u_i\|^2 | \sigma^2] \leq \xi_i \sigma_e^2 + \zeta_i \sum_{j=1}^r \sigma_{u_j}^2 + \text{const} .$$

A.4 An upper bound on $E[(\|u_i\|^2)^{-c} | \sigma^2]$

Fix $i \in \{1, 2, \dots, r\}$. Given σ^2 , $(R_i Q^{-1} R_i^T)^{-1/2} u_i$ has a multivariate normal distribution with identity covariance matrix. It follows that, conditional on σ^2 , the distribution of $u_i^T (R_i Q^{-1} R_i^T)^{-1} u_i$ is $\chi_{d_i}^2(w)$. It follows from Lemma 4 that, as long as $c \in (0, 1/2)$, we have

$$E[u_i^T (R_i Q^{-1} R_i^T)^{-1} u_i | \sigma^2]^{-c} \leq 2^{-c} \frac{\Gamma(\frac{q_i}{2} - c)}{\Gamma(\frac{q_i}{2})} .$$

Now, by Lemma 1

$$\begin{aligned}
E\left[\left(\|u_i\|^2\right)^{-c} \mid \sigma^2\right] &= \left((\sigma_e^2)^{-1} \lambda_{\max} + (\sigma_{u_i}^2)^{-1}\right)^c E\left[\left[u_i^T \left((\sigma_e^2)^{-1} \lambda_{\max} + (\sigma_{u_i}^2)^{-1}\right) I_{q_i} u_i\right]^{-c} \mid \sigma^2\right] \\
&\leq \left((\sigma_e^2)^{-1} \lambda_{\max} + (\sigma_{u_i}^2)^{-1}\right)^c E\left[\left[u_i^T (R_i Q^{-1} R_i^T)^{-1} u_i\right]^{-c} \mid \sigma^2\right] \\
&\leq \left((\sigma_e^2)^{-1} \lambda_{\max} + (\sigma_{u_i}^2)^{-1}\right)^c 2^{-c} \frac{\Gamma\left(\frac{q_i}{2} - c\right)}{\Gamma\left(\frac{q_i}{2}\right)} \\
&\leq 2^{-c} \frac{\Gamma\left(\frac{q_i}{2} - c\right)}{\Gamma\left(\frac{q_i}{2}\right)} \left[\lambda_{\max}^c (\sigma_e^2)^{-c} + (\sigma_{u_i}^2)^{-c}\right].
\end{aligned}$$

References

- DANIELS, M. J. (1999). A prior for the variance in hierarchical models. *Canadian Journal of Statistics*, **27** 569–580.
- DIACONIS, P., KHARE, K. and SALOFF-COSTE, L. (2008). Gibbs sampling, exponential families and orthogonal polynomials (with discussion). *Statistical Science* 151–200.
- FLEGAL, J. M., HARAN, M. and JONES, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science*, **23** 250–260.
- GELFAND, A. E., SAHU, S. K. and CARLIN, B. P. (1995). Efficient parametrisations for normal linear mixed models. *Biometrika*, **82** 479–488.
- GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **1** 515–534.
- JOHNSON, A. A. and JONES, G. L. (2010). Gibbs sampling for a Bayesian hierarchical general linear model. *Electronic Journal of Statistics*, **4** 313–333.
- JONES, G. L., HARAN, M., CAFFO, B. S. and NEATH, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association*, **101** 1537–1547.
- KHARE, K. and HOBERT, J. P. (2011). A spectral analytic comparison of trace-class data augmentation algorithms and their sandwich variants. *Annals of Statistics* (to appear).
- LIU, J. S., WONG, W. H. and KONG, A. (1994). Covariance structure of the Gibbs sampler with applications to comparisons of estimators and augmentation schemes. *Biometrika*, **81** 27–40.

- MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer Verlag, London.
- PAPASPILIOPOULOS, O. and ROBERTS, G. O. (2008). Stability of the Gibbs sampler for Bayesian hierarchical models. *Annals of Statistics*, **36** 95–117.
- PAPASPILIOPOULOS, O., ROBERTS, G. O. and SKÖLD, M. (2007). A general framework for the parametrization of hierarchical models. *Statistical Science*, **22** 59–73.
- ROBERTS, G. O. and ROSENTHAL, J. S. (1998). Markov chain Monte Carlo: Some practical implications of theoretical results (with discussion). *Canadian Journal of Statistics*, **26** 5–31.
- ROBERTS, G. O. and ROSENTHAL, J. S. (2001). Markov chains and de-initializing processes. *Scandinavian Journal of Statistics*, **28** 489–504.
- ROBERTS, G. O. and ROSENTHAL, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probability Surveys*, **1** 20–71.
- ROMÁN, J. C. (2012). *Convergence Analysis of Block Gibbs Samplers for Bayesian General Linear Mixed Models*. Ph.D. thesis, Department of Statistics, University of Florida.
- SEARLE, S. R., CASELLA, G. and MCCULLOCH, C. E. (1992). *Variance Components*. John Wiley & Sons, New York.
- SUN, D., TSUTAKAWA, R. and HE, Z. (2001). Propriety of posteriors with improper priors in hierarchical linear mixed models. *Statistica Sinica*, **11** 77–95.
- TAN, A. and HOBERT, J. P. (2009). Block Gibbs sampling for Bayesian random effects models with improper priors: Convergence and regeneration. *Journal of Computational and Graphical Statistics*, **18** 861–878.
- YU, Y. and MENG, X.-L. (2011). To center or not to center: That is not the question - An ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC efficiency (with discussion). *Journal of Computational and Graphical Statistics*, **20** 531–570.